

A Feature Weighing-Assisted Approach for Cancer Subtypes Identification from Paired Expression Profiles

Madhumita and Sushmita Paul

Department of Bioscience & Bioengineering,
Indian Institute of Technology, Jodhpur

NH. 65, Surpura Bypass Rd, Karwar, Jodhpur, Rajasthan 342037, India
{[madhumita.1](mailto:madhumita.1@iitj.ac.in), [sushmitapaul](mailto:sushmitapaul@iitj.ac.in)} @iitj.ac.in



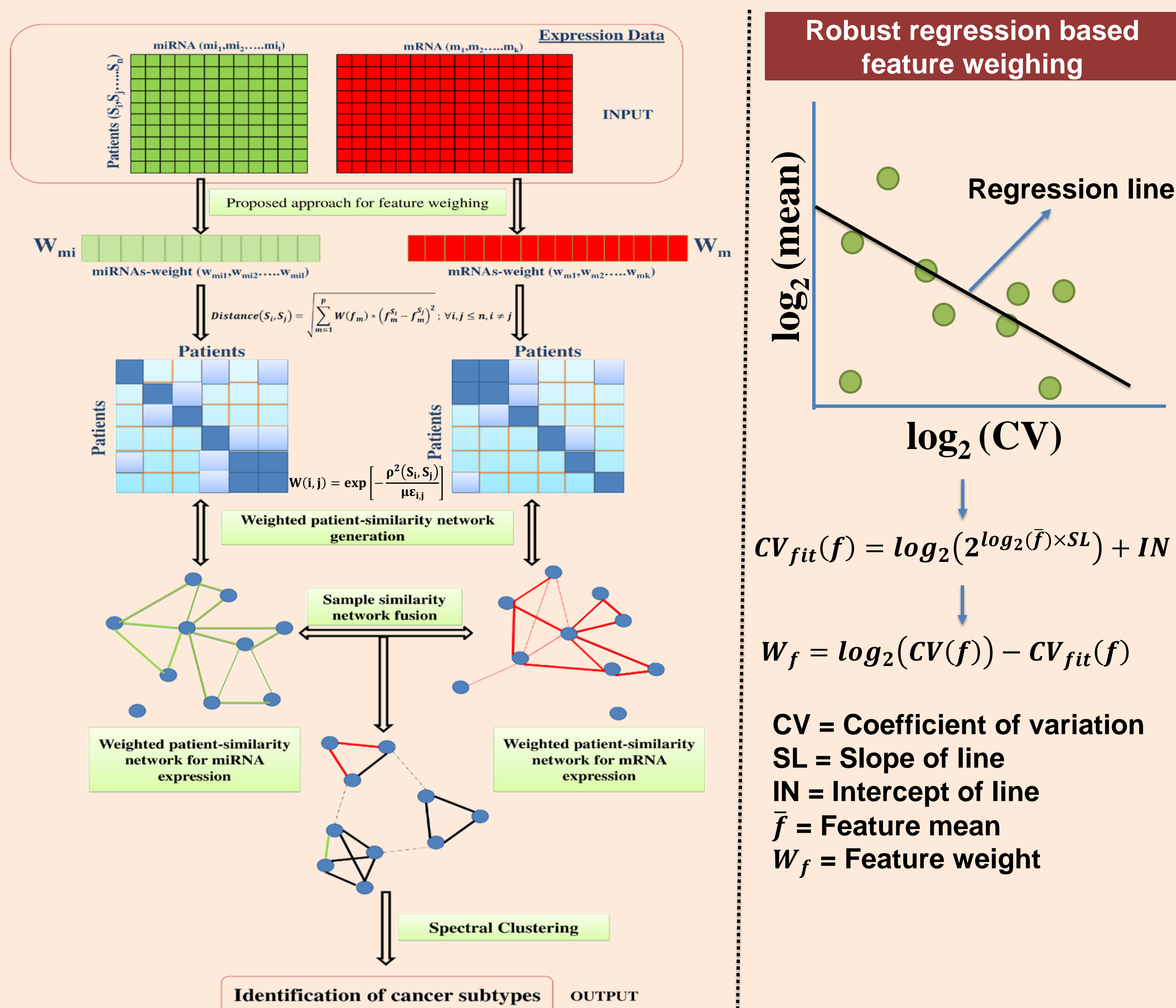
॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

Department of Bioscience & Bioengineering

Introduction

- Cancer-subtyping is critically important for understanding the heterogeneity present in tumours.
- Availability of expression profiles of multiple types of bio-markers across same set of samples can help in capturing this heterogeneity underlying, complex biological processes and phenotypes.
- Robust-regression based feature ranking, prior to the integration of multiple information sources (miRNA and mRNA expression profiles), may help in identification of histological subtypes of solid tumours.
- This can further help in analysing shared molecular profiles between different groups for designing appropriate therapies and treatments.

Proposed Approach for Identification of Histological Subtypes within a Cancer type



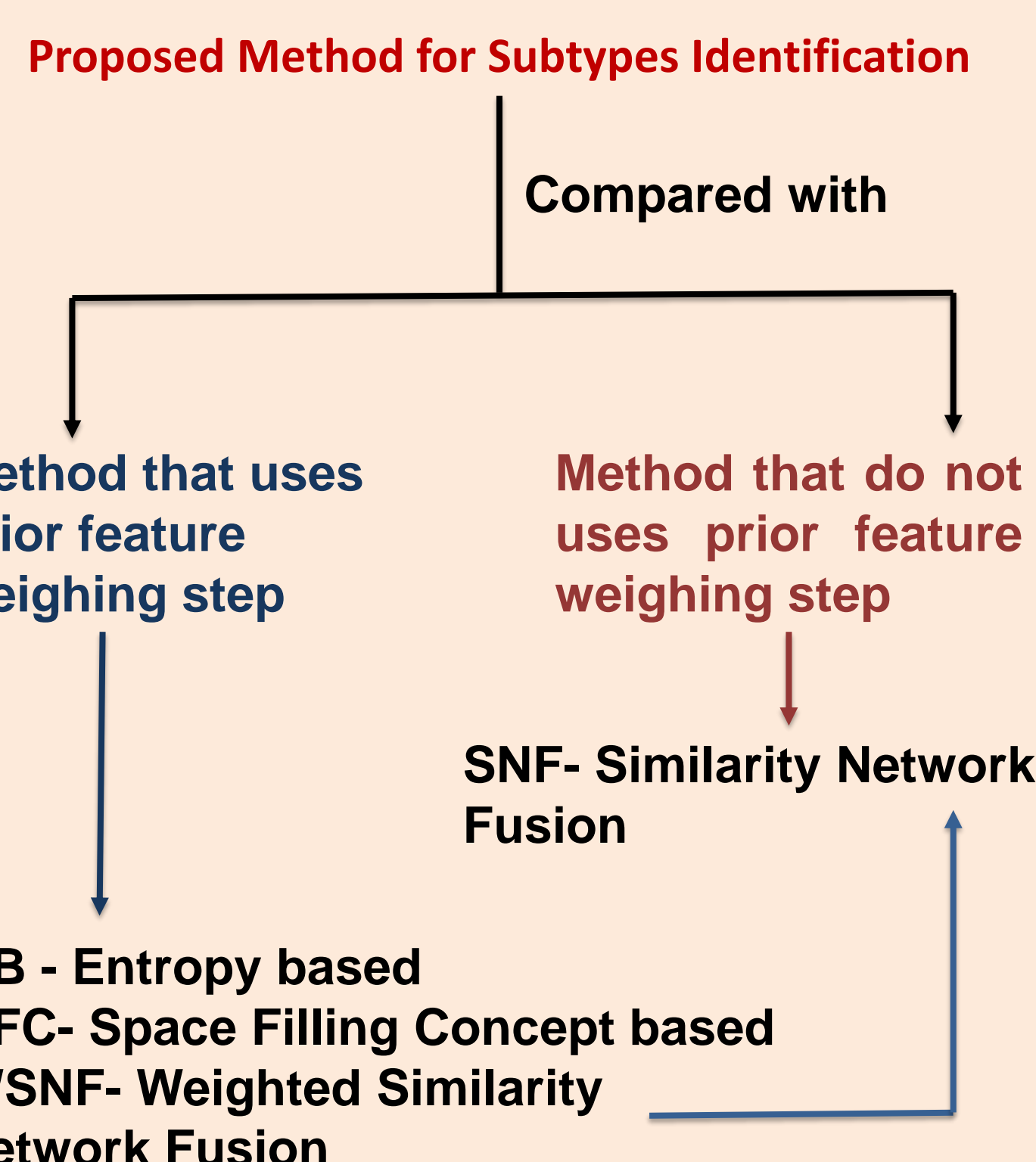
Experimental Setup

Data sets used

- ✓ TCGA data sets used here are freely available at <http://www.cbioportal.org/>
- ✓ Each cancer has 2 histological subtypes.

Cervical Cancer
Adenocarcinoma
Squamous Cell Carcinoma
Breast Cancer
Infiltrating Ductal Carcinoma
Infiltrating Lobular Carcinoma
Ovarian Cancer
Neoplasm Histological Grade 2
Neoplasm Histological Grade 3

Method Comparison



Results

Clustering Results Comparison

Cancer types	Cluster evaluating indices	Proposed	EB	SFC	WSNF
Cervical Cancer	Average Silhouette Width	0.48	0.39	0.37	0.01
	Accuracy	0.89	0.82	0.80	0.63
	Purity	0.75	0.70	0.68	0.58
	Normalized Mutual Information	0.39	0.25	0.21	0.19
Breast Cancer	Average Silhouette Width	0.84	0.80	0.80	0.81
	Accuracy	0.59	0.54	0.54	0.54
	Purity	0.60	0.53	0.53	0.52
	Normalized Mutual Information	0.19	0.18	0.04	0.18
Ovarian Cancer	Average Silhouette Width	0.63	0.48	0.47	0.51
	Accuracy	0.74	0.64	0.64	0.46
	Purity	0.73	0.60	0.60	0.44
	Normalized Mutual Information	0.22	0.19	0.00	0.19

Differential Expression Analysis Between Subtypes of Cancer

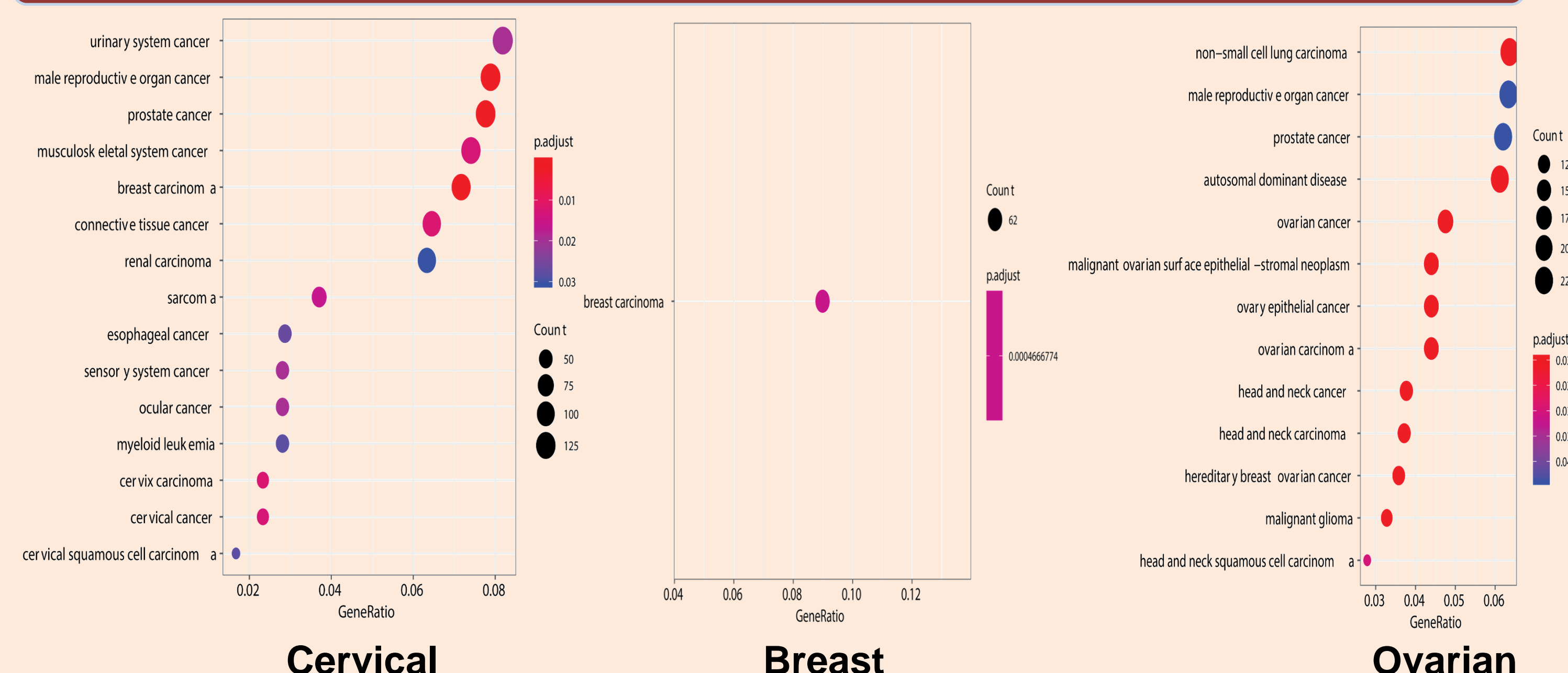


Fig 1: **Disease Ontology (DO) Enrichment** : DO plots for the set of differentially expressed genes between cancer subtypes of respective cancers identified by the proposed method

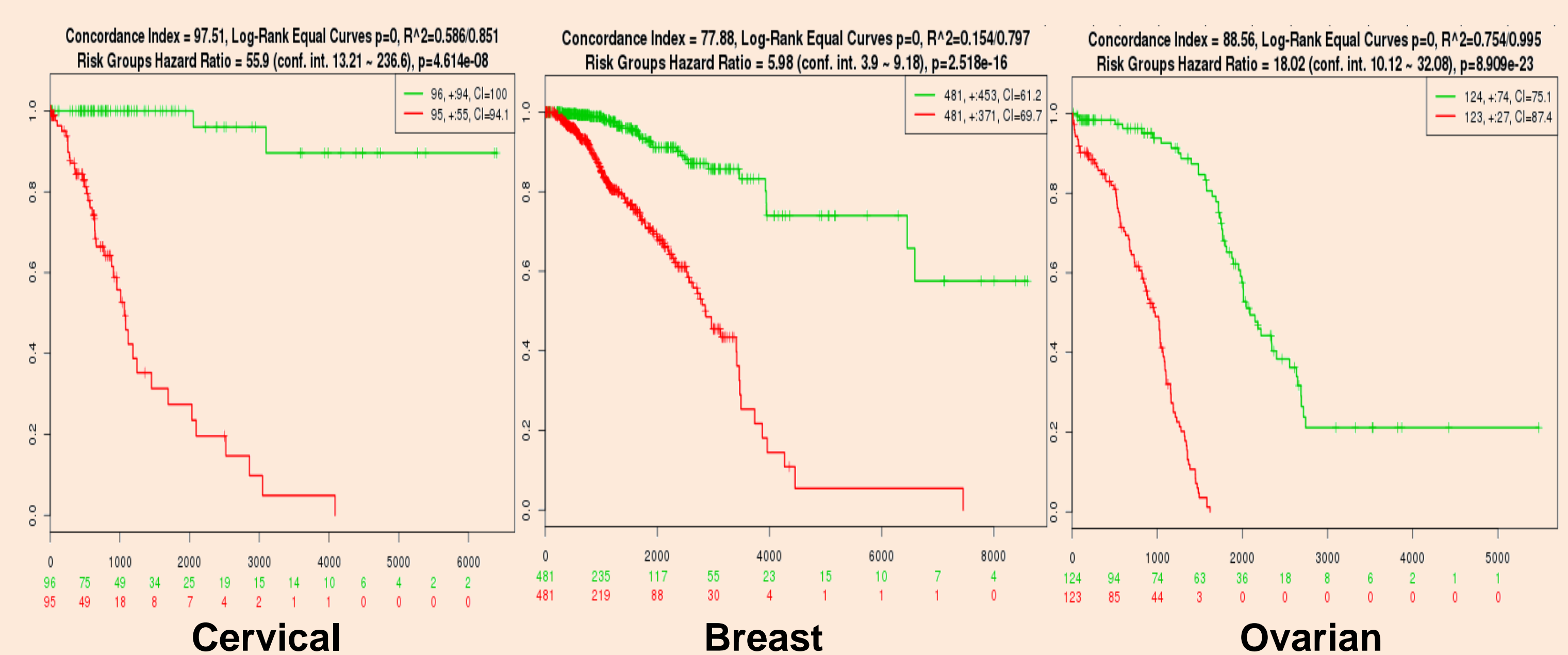


Fig 2: **Survival Analysis** : Kaplan-Meier curves for cancer patients plotted for combined expression of top 100 differentially expressed genes between respective subtypes obtained by the proposed approach

Conclusion

- Application of robust regression based feature weighing method, prior to the multiple information fusion and clustering, improves the clustering efficiency.
- The feature weight assigned to every biomarker, efficiently captures the variation in their expression across the samples. This helps in exploring the dynamic nature of the expression data measured under different biological condition.
- Differential expression of the biomarkers between the identified histological subtypes also shows biological relevance and can be further explored for getting deeper understanding about the molecular profile.

Acknowledgment

This work is partially supported by the seed grant program of the Indian Institute of technology Jodhpur, India (grant no.- I/SEED/SPU/20160010)